R E P O R T R E S U M E S

ED 013 280 UD 004 020

FACTOR ANALYSES OF ACHIEVEMENT MEASURES FROM THE EDUCATIONAL OPPORTUNITIES SURVEY.

BY- MAYESKE, GEORGE W. WEINFELD, FREDERIC D.
NATIONAL CENTER FOR EDUCATIONAL STATISTICS ØDHEWØ

REPORT NUMBER TN-21 PUB DATE 18 JAN 67
EDRS PRICE MF-$0.25 HC-$0.84 21P.

DESCRIPTORS- *FACTOR ANALYSIS, *ACHIEVEMENT, VERBAL ABILITY, READING COMPREHENSION, MATHEMATICS, GRADE 9, MEASUREMENT TECHNIQUES, EDUCATIONAL OPPORTUNITIES SURVEY

FACTOR ANALYSES WERE APPLIED TO TABLES OF INTERCORRELATIONS ORIGINALLY COMPUTED FOR 11 GROUPS OF NINTH GRADERS (NORTHERN AND SOUTHERN NEGRO AND WHITE, PUERTO RICAN, MEXICAN, INDIAN, AND ORIENTAL-AMERICAN) ON FIVE ACHIEVEMENT MEASURES USED BY COLEMAN AND OTHERS IN THE EDUCATIONAL OPPORTUNITIES SURVEY. THE PURPOSE OF THE FACTOR ANALYSIS WAS TO DETERMINE WHETHER THE FIVE TESTS--NONVERBAL ABILITY, VERBAL ABILITY, READING COMPREHENSION, MATHEMATICS ACHIEVEMENT, AND GENERAL INFORMATION--HAD ENOUGH IN COMMON IN WHAT THEY WERE MEASURING TO BE COMBINED INTO ONE SCORE--AN "INDEX OF ACHIEVEMENT" SCORE. IN ORDER TO MEET THE REQUIREMENTS FOR COMBINING THE FIVE SCORES INTO ONE, TWO REQUIREMENTS OF THE FACTOR ANALYSIS WERE ESSENTIAL--(1) THAT THE FIVE MEASURES WERE MEASURING TO A HIGH DEGREE ONLY ONE THING IN COMMON (WHICH WOULD BE SHOWN IF THE FACTOR ANALYSIS FOUND HIGH VALUES FOR THE FIRST FACTOR EXTRACTED), AND (2) THAT WHEN "FIRST FACTORS" WERE COMPUTED FOR EACH OF 10 SUBGROUPS, THESE FIRST FACTORS WOULD NOT DEVIATE FROM THE FIRST FACTOR EXTRACTED FROM THE TOTAL GROUP. THE RESULTS OF 11 FACTOR ANALYSES INDICATED THAT BOTH OF THESE REQUIREMENTS WERE MET, AND THEREFORE THE NUMBERS OF THE FIRST FACTOR OF THE TOTAL GROUP COULD BE EMPLOYED AS "WEIGHTS" TO MULTIPLY EACH INDIVIDUAL'S SCORE ON EACH OF THE FIVE MEASURES FROM THE COLEMAN STUDY SO THAT THEY MIGHT BE COMBINED (ADDED) TO FORM ONE INDEX OF ACHIEVEMENT SCORE. THE WEIGHTS FOR THE FIVE TESTS ARE--NON-VERBAL (.76), VERBAL (.92), READING COMPREHENSION (.87), MATHEMATICAL ACHIEVEMENT (.85), AND GENERAL INFORMATION (.91). (WT)

ED013280

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Working Paper
Does Not Reflect
Official Policy of
The Office of Education

UD004020

NATIONAL CENTER FOR EDUCATIONAL STATISTICS
Division of Operations Analysis

FACTOR ANALYSES OF ACHIEVEMENT MEASURES FROM THE EDUCATIONAL OPPORTUNITIES SURVEY

by

George W. Mayeske
Frederic D. Weinfeld

Technical Note
Number 21

January 18, 1967

OFFICE OF EDUCATION/U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

NATIONAL CENTER FOR EDUCATIONAL STATISTICS
Alexander M. Mood, Assistant Commissioner

DIVISION OF OPERATIONS ANALYSIS
David S. Stoller, Director

TABLE OF CONTENTS

	Page
List of Tables	4 & 6, 10-15
Statement of the Problem	1
Description of Analytical Techniques	1
The Relationship of the First Principal Component to the First Centroid of a Correlation Matrix with Positive Manifold	2
Analyses and Results by Racial, Ethnic and Regional Groups	3
List of References Cited	8
Appendix A-Group Intercorrelations, Means, and Standard Deviations	9
Appendix B-Factor Iterations and Convergence Criteria for the First Principal Component	16

FACTOR ANALYSES OF ACHIEVEMENT MEASURES FROM THE EDUCATIONAL OPPORTUNITIES SURVEY

Statement of the Problem

The recently completed Educational Opportunities Survey (Coleman et al. 1966, see page 8 for the full reference) obtained a large number of measures of the attributes of students, their home backgrounds, their teachers, principals and superintendents, and of their community. The objective of current analyses of the Educational Opportunities Survey (hereafter referred to as (EOS) data) is to reduce this large number of measures to a smaller number of basic indexes and then conduct complex analyses of these indexes to determine which combinations facilitate educational achievement. Accordingly, it was first desired to obtain an index of achievement from the five EOS achievement measures.

Factor analysis was considered to be a useful tool for index development. This technique analyzes what is common to a set of measures and attempts to reduce them to a more basic and less numerous set of factors. The commonness among a set of measures is usually expressed by the use of covariances or correlation coeffecients. When the correlations are high the measures are said to assess similar or common attributes. A factor is a hypothetical variable which is introduced to account for these correlations. The end result of a factor analysis is a set of one or more factors where each measure has a weight (or loading) for each factor to indicate the extent of its relationship to that factor. These factors when multiplied together, will reproduce the correlation coefficients to some desired degree of accuracy. The weights or loadings can be used to combine the measures into an index by multiplying each person's score on each measure by the weight for that measure. The five measures from the EOS are:

1. Non-Verbal Ability
2. Verbal Ability
3. Reading Comprehension
4. Mathematics Achievement
5. General Information Total

The reader is referred to the EOS report (Coleman et al. 1966) for a full description of these measures and examples of items from each.

Description of the Analytical Techniques

The intercorrelations of the above measures for different ethnic, racial and regional groups were available from the published EOS data and files. In order to expedite the development of the achievement index which was needed for subsequent analyses it was decided to compute the factors using a desk calculator.

The analyses were primarily concerned with what the measures had in common. Thus, the relation of each measure with the first factor (commonly called factor loadings) was of major interest. If the measures all related (loaded) highly with the first factor then a single achievement index would suffice. If some measures had low loadings with the first factor then a second factor and hence a second achievement index might be required. In addition, if the measures relate quite differently with the first factor for different racial or regional groups then the same set of weights could not be applied to the measures for these different groups to obtain an achievement index. In this case a slightly different technique would be required.

When a computer is readily available the principal component method of factor extraction is the most preferred technique. This is so because this technique has the most desirable mathematical properties, that is, it maximizes the amount of variance taken out with each factor and hence takes out the largest factor first, then the next largest, etc. For desk calculator techniques the centroid method of factor extraction is widely used as an approximation to the principal component method because it is computationally simpler.

<u>The Relationship of the First Principal Component to the First Centroid of a Correlation Matrix with Positive Manifold</u>

Let R be an n th order symmetric matrix of correlation coefficients. Positive manifold is said to exist when all of the coefficients are non-negative. Under these conditions we are interested in using the first centroid of the correlation matrix as an approximation to the first principal component. The following comparisons are based upon derivations taken from Horst's book entitled "The Factor Analysis of Data Matrices" (see page 8).

Let U be an n th order column vector with unit elements. Then the first centroid factor is given by:

(1) $F_c = \frac{RU}{\sqrt{U^T RU}}$ where U^T designates the transpose of U.

The principal component method extracts the maximum amount of variance for each successive factor by iterating a solution for each factor until the magnitude of the factor weights reach a plateau, that is, until the factor weights for each measure do not differ by more than a small prespecified amount from one iteration to the next. The formula for the i th iteration is:

(2) $F_h = \frac{R^i V}{\sqrt{V^T R^{2i-1} V}}$ where V denotes an arbitrary vector of initial weights.

Since V is arbitrary, let it equal U, the unit column vector. Then the formula becomes:

$$F_h = \frac{R^i U}{\sqrt{U^T R^{2i-1} U}} \qquad (3)$$

Clearly, for the initial iteration (1) and (3) are identical. For further iterations they differ according to differences in higher powers of the R matrix. Empirically, the weights for the measures tend to be higher for the principal component than for the centroid method. However, this effect is not uniform for each measure.

Analyses and Results by Racial, Ethnic and Regional Groups

Centroid factors were computed for the following ninth grade groups:

1. Total ninth grade (TN)
2. Mexican-American (MA)
3. Puerto-Ricans (PR)
4. Indian-Americans (IA)
5. Oriental-Americans (OA)
6. Total Negro ninth grade (N total)
7. Total White ninth grade (W total)
8. Negro North (N North)
9. Negro South (N South)
10. White North (W North)
11. White South (W South)

The means, standard deviations, intercorrelations and first centroids for each group are given in Tables 3 through 13 in Appendix A. The data for groups 2 through 11 were obtained from the EOS Supplemental Appendix. The data for group 1 was taken from EOS files.

In addition to the first centroid the first principal component was computed for group 1 to check on the accuracy of the centroid as an approximation to it. These computations are given in Appendix B. The first principal component converged after two iterations. It is presented in Table 1 with the centroid factors for each group.

Inspection of the first two columns of Table 1 shows that the first principal component and the first centroid are almost identical. The main difference is in the non-verbal measure which is very slightly more highly related to the centroid than the principal component. Thus the first centroid has given a very close approximation.

TABLE 1

Loadings of Each Measure with the
First Factor of Each Group

	TOTAL											
	Prin. Comp.	Centroid	MA	PR	IA	OA	N TOTAL	W TOTAL	N NORTH	N SOUTH	W NORTH	W SOUTH
Non-verbal 1	.76	.79	.80	.75	.73	.81	.75	.75	.72	.76	.75	.76
Verbal 2	.92	.91	.88	.87	.87	.88	.88	.89	.87	.88	.89	.89
Readg. Comp. 3	.87	.87	.85	.85	.86	.86	.83	.85	.82	.83	.84	.85
Math. Ach. 4	.85	.85	.82	.78	.79	.86	.76	.84	.77	.75	.84	.83
Gen'l.Info. 5	.91	.90	.88	.85	.88	.89	.86	.88	.86	.85	.88	.88

The maximum value that a loading or weight can have is 1.00. Table 1 shows that the measures all have high loadings on the first factor. Hence a single index of achievement is all that is required to represent the inter-relationships among the achievement measures.

Inspection of the other group factors shows that the relative ordering of the measures on each factor as well as their absolute level are highly consistent. The relative pattern in descending order with only slight deviations, is: Verbal; General Information; Reading Comprehension; Mathematics; and Non-verbal. The absolute deviation of the principal component factor loading from the centroid factor loading for each group is given in Table 2. These values are obtained by taking the difference, without regard to sign, between the loadings in column 1 and each of the remaining columns (starting with column MA) in Table 1. Thus the entry for the first row and column in Table 2 is obtained by taking the difference between .76 from row 1 and column 1 of Table 1 and .80 from row 1 and column 3 of Table 1.

TABLE 2

Absolute Deviation of Each Group
Factor from the Principal Component for the Total Group

		MA	PR	IA	OA	N TOTAL	W TOTAL	N NORTH	N SOUTH	W NORTH	W SOUTH	ROW AVERAGE
Non-verbal	1	.04	.01	.03	.05	.01	.01	.04	.00	.01	.00	.020
Verbal	2	.04	.03	.05	.04	.04	.03	.05	.04	.03	.03	.038
Readg. Comp.	3	.02	.02	.01	.01	.04	.02	.05	.04	.03	.02	.026
Math. Ach.	4	.03	.07	.06	.01	.09	.01	.08	.01	.01	.02	.048
Gen'l. Info.	5	.03	.06	.03	.03	.05	.03	.05	.06	.03	.03	.040
Group Average		.032	.038	.036	.028	.046	.020	.054	.048	.022	.020	.0344

Mathematics Achievement shows the greatest average deviation and General Information the next largest deviation from the principal component. The groups that show the greatest average deviation are Northern and Southern Negroes. These decrepancies are determined largely by Mathematics Achievement which tends to load slightly less with the first factor for these groups than for the others. These differences are not appreciable and do not alter the relative ordering of the measures on the factor. Hence, a single weighting system can be used to obtain an achievement index for these different groups.

Therefore, it is decided to use the factor loadings from the first principal component of the "Total" matrix as the weights to obtain this index. These weights are:

Non-verbal	.76
Verbal	.92
Readg. Comp.	.87
Math. Ach.	.85
Gen'l. Info.	.91

LIST OF REFERENCES CITED

Coleman, James S., et al. Equality of Educational Opportunity. National Center for Educational Statistics, U.S. Government Printing Office, Washington: 1966, Catalog No. FS5. 38001 and Supplement.

Horst, Paul, Factor Analysis of Data Matrices, Holt, Rhinehart and Winston, Inc.: 1965.

APPENDIX A

Group Means*, Standard Deviations and Intercorrelations

*The means for measures 1 and 5 are in raw score form whereas those for measures 2, 3 and 4 are scale values.

TABLE 3

Intercorrelations, Means, and Standard Deviations
for the Total Group Grade 9

	1	2	3	4	5	MEAN	STD. DEV.
1	1.0000	.6039	.5900	.6158	.5955	36.3820	8.9923
2	.6039	1.0000	.7670	.6894	.8500	50.6914	15.2527
3*	.5900	.7670	1.0000	.6522	.7282	57.1647	18.4632
4*	.6158	.6894	.6522	1.0000	.7058	45.0229	15.1320
5*	.5955	.8500	.7282	.7058	1.0000	50.0566	15.8975

TABLE 4

Intercorrelations, Means, and Standard Deviations
for Mexican-Americans Grade 9

	1	2	3	4	5	MEAN	STD. DEV.
1	1.0000	.6121	.5921	.5914	.6015	31.7365	10.0525
2	.6121	1.0000	.7111	.6237	.7918	41.9519	13.5295
3*	.5921	.7111	1.0000	.6119	.6807	47.7315	18.1208
4*	.5914	.6237	.6119	1.0000	.6395	37.0561	15.5670
5*	.6015	.7918	.6807	.6395	1.0000	40.2244	13.9311

*Measures 3, 4 and 5 appear as variable numbers.

TABLE 5

Intercorrelations, Means, and Standard Deviations
for Puerto Ricans Grade 9

	1	2	3	4	5	MEAN	STD. DEV.
1	1.0000	.5311	.5687	.5262	.4592	30.6022	10.2599
2	.5311	1.0000	.7163	.5454	.7797	38.6336	31.8322
3	.5687	.7163	1.0000	.5504	.6611	42.6623	17.2668
4	.5262	.5454	.5504	1.0000	.5824	32.7753	15.1201
5	.4592	.7797	.6611	.5824	1.0000	35.0150	13.1499

TABLE 6

Intercorrelations, Means, and Standard Deviations
for Indian-Americans Grade 9

	1	2	3	4	5	MEAN	STD. DEV.
1	1.0000	.4788	.5401	.4928	.4869	33.9188	9.3535
2	.4788	1.0000	.7260	.5834	.8092	43.3427	13.7061
3	.5401	.7260	1.0000	.5732	.7131	49.2725	16.9029
4	.4928	.5834	.5732	1.0000	.6126	38.1914	14.0336
5	.4869	.8092	.7131	.6126	1.0000	41.8858	14.4126

TABLE 7

Intercorrelations, Means, and Standard Deviations
for Oriental-Americans Grade 9

	1	2	3	4	5	MEAN	STD. DEV.
1	1.0000	.5698	.6186	.6693	.6203	38.2663	9.4696
2	.5698	1.0000	.7141	.6767	.8295	50.4940	14.4194
3	.6186	.7141	1.0000	.6802	.6789	57.6727	18.4471
4	.6693	.6767	.6802	1.0000	.6860	47.6386	16.2217
5	.6203	.8295	.6789	.6860	1.0000	50.3363	15.1100

TABLE 8

Intercorrelations, Means, and Standard Deviations for
Negro Total Grade 9

	1	2	3	4	5	MEAN	STD. DEV.
1	1.0000	.5402	.5102	.4952	.5139	30.2161	9.7906
2	.5402	1.0000	.7002	.5483	.7905	38.1583	12.5450
3	.5102	.7002	1.0000	.5143	.6656	44.5471	16.4791
4	.4952	.5483	.5143	1.0000	.5519	34.5399	13.8912
5	.5139	.7905	.6656	.5519	1.0000	36.8132	12.3747

TABLE 9

Intercorrelations, Means, and Standard Deviations for
White Total Grade 9

	1	2	3	4	5	MEAN	STD. DEV.
1	1.0000	.5360	.5307	.5760	.5290	38.1820	7.8596
2	.5360	1.0000	.7321	.6526	.8210	54.4202	14.0359
3	.5307	.7321	1.0000	.6143	.6807	61.0122	17.2630
4	.5760	.6526	.6143	1.0000	.6724	48.2141	13.9261
5	.5290	.8210	.6807	.6724	1.0000	54.0855	14.6347

TABLE 10

Intercorrelations, Means, and Standard Deviations for
Negro North Grade 9

	1	2	3	4	5	MEAN	STD. DEV.
1	1.0000	.4808	.4702	.4863	.4779	32.5059	9.1105
2	.4808	1.0000	.6760	.5542	.7876	40.8533	12.5671
3	.4702	.6760	1.0000	.5114	.6541	47.0518	16.4675
4	.4863	.5542	.5114	1.0000	.5671	36.0067	13.7755
5	.4779	.7876	.6541	.5671	1.0000	38.9957	12.8352

TABLE 11

Intercorrelations, Means, and Standard Deviations
for Negro South Grade 9

	1	2	3	4	5	MEAN	STD. DEV.
1	1.0000	.5526	.5177	.4877	.5152	27.9363	9.9114
2	.5526	1.0000	.7084	.5281	.7782	35.4750	11.9328
3	.5177	.7084	1.0000	.5026	.6609	42.0534	16.1083
4	.4877	.5281	.5026	1.0000	.5221	33.0794	13.8250
5	.5152	.7782	.6609	.5221	1.0000	34.6403	11.4939

TABLE 12

Intercorrelations, Means, and Standard Deviations
for White North Grade 9

	1	2	3	4	5	MEAN	STD. DEV.
1	1.0000	.5204	.5216	.5754	.5160	38.9167	7.4753
2	.5204	1.0000	.7296	.6567	.8235	55.4182	13.8927
3	.5216	.7296	1.0000	.6094	.6763	61.7639	17.1962
4	.5754	.6567	.6094	1.0000	.6749	49.0418	13.9712
5	.5160	.8235	.6763	.6794	1.0000	55.0557	14.6827

TABLE 13

Intercorrelations, Means, and Standard Deviations
for White South Grade 9

	1	2	3	4	5	MEAN	STD. DEV.
1	1.0000	.5485	.5426	.5668	.5406	36.4978	8.4385
2	.5485	1.0000	.7336	.6326	.8092	52.1322	14.0947
3	.5426	.7336	1.0000	.6191	.6856	59.2891	17.2922
4	.5668	.6326	.6191	1.0000	.6562	46.3165	13.6339
5	.5406	.8092	.6856	.6562	1.0000	51.8615	14.2775

APPENDIX B

Factor Iterations and Convergence Criteria for the First Principal Component.